

GERAÇÃO DE CONJUNTO DE DADOS SINTÉTICO PARA DETECÇÃO DE INSTÂNCIAS DE OBJETOS

Bruno de Freitas Vece Perez¹, Plínio Thomaz Aquino Junior²

¹ Departamento de Engenharia Elétrica, Centro Universitário FEI

² Departamento de Ciência da Computação, Centro Universitário FEI
uniebperez@fei.edu.br, plinio.aquino@fei.edu.br

Resumo: Atualmente robôs estão sendo utilizados em ambientes domésticos para executar tarefas como limpar a casa e monitorar o ambiente. Esse projeto considera a pesquisa e desenvolvimento de uma abordagem para geração de uma base de dados, a partir de uma quantidade restrita de imagens coletadas dos objetos de interesse. O principal objetivo é desenvolver uma ferramenta para criação de um conjunto de dados, de forma rápida, para treinamento de Rede de detecção de objetos, a partir da geração de imagens sintéticas.

1. Introdução

Um robô de serviço deve conseguir reconhecer de maneira específica os objetos contidos nos ambientes em que ele atua. O surgimento da aprendizagem profunda e seu uso em visão computacional, considerando os resultados surpreendentes que as redes neurais profundas estavam gerando, fez com que pesquisadores começassem a implementar soluções de aprendizagem profunda para resolver problemas de detecção de objetos.

A equipe RoboFEI@Home, envolvida nesse projeto, desenvolve um robô de serviço que deve ser capaz de reconhecer os objetos inseridos no ambiente em que ele atua. Segundo [1], o método de detecção de objetos desenvolvido pela equipe é baseado no uso de rede neural convolucional. Com esta abordagem, um conjunto de dados contendo fotos de objetos domésticos são necessárias. Exemplos de objetos devem ser marcado manualmente com caixas delimitadoras para gerar um conjunto de dados consistente.

Embora os métodos usando aprendizado profundo sejam eficazes na solução de problemas envolvendo detecção de objetos, seu funcionamento depende de uma grande quantidade de dados coletados e anotados manualmente. Esse não é um problema caso deseja-se fazer uma detecção mais genérica dos objetos contidos no ambiente, pois, para isso, é possível encontrar bases de dados em larga escala com classes mais genéricas de objetos. Porém, quando deseja-se fazer detecções específicas de instâncias de objetos, imagens dos objetos precisam ser coletadas manualmente e cada objeto deve ser marcado nas imagens de treinamento com caixas delimitadoras, sendo esse processo muito demorado.

Em alguns casos, adicionar dados gerados sinteticamente pode ser uma boa maneira de fazer o aumento dos dados de um *dataset* e melhorar o desempenho, gastando menos tempo para geração do conjunto de dados. Alguns trabalhos focam no uso das redes geradoras adversárias (GAN, do inglês *Generative Adversarial Networks*), como é o caso de [2], que usa modelos 3D dos objetos de interesse e um mecanismo de simulação para inclui-los em um ambiente simulado e usar uma GAN para tornar as imagens mais realistas.

Outros trabalhos tentam ir por um caminho mais simples, como é o caso de [3]. Em seu trabalho, criaram um algoritmo que consiste em gerar máscaras dos objetos que se deseja detectar usando a segmentação de imagem e, em seguida, sobrepor e mesclar o recorte desses objetos com diferentes imagens de fundo. Segundo os autores, o método supera as abordagens de síntese existentes e, quando combinado com imagens reais, melhora o desempenho relativo dos detectores em mais de 21%.

Desta forma, esse trabalho tem como objetivo desenvolver uma ferramenta semelhante à descrita no trabalho citado anteriormente. Porém, será utilizada uma abordagem diferente para a geração das máscaras dos objetos e composição das novas imagens. O projeto foi dividido em duas partes principais, a primeira é um método para extrair automaticamente o fundo das imagens de amostra dos objetos usando *Deep Salient Object Detection* [4] e a segunda é um método para usar essas imagens e as máscaras geradas para criar novas imagens.

2. Metodologia

Para a etapa de segmentação do objeto, com o objetivo de extração do fundo, utilizou-se a Detecção de Objetos Proeminentes, por ser a opção que apresentou bons resultados. Foram realizados alguns testes com alguns algoritmos de visão computacional para remoção de fundo, além de alguns testes com tela verde, porém os resultados obtidos não foram satisfatórios como o método apresentado. Foi decidido que seria utilizada a arquitetura proposta por [4], por ter demonstrado um bom desempenho nas diversas avaliações comparativas realizadas. Os autores estenderam a arquitetura HED [5] para atingir o objetivo da detecção de saliências. Como a detecção de saliências é relativamente mais complicada que a detecção de bordas, eles adicionaram mais duas camadas convolucionais a rede. Para implementar a arquitetura de rede neural proposta, foi utilizado o Keras, uma biblioteca para aprendizado profundo desenvolvido em Python, que roda em cima do TensorFlow [6].

O conjunto de dados utilizado para o treinamento do modelo foi o MSRA-B [7], um conjunto de dados que contém cerca de 5000 imagens de centenas de categorias diferentes. Por sua diversidade e grande quantidade de imagens disponíveis, esse *dataset* tem sido um dos conjuntos de dados mais usados na literatura que está disponível publicamente.

Após a geração das máscaras, tem-se a etapa de geração das imagens finais. Para essa etapa, foi desenvolvido um programa Python que utiliza a imagem do objeto e sua máscara binária, junto com as imagens de fundo, para compor as imagens finais. Na geração é possível definir algumas propriedades como o número de

vezes que um objeto aparecerá no *dataset*, rotação máxima dos objetos na colagem, redimensionamento máximo e mínimo das imagens dos objetos, número de objetos por imagem final e outras propriedades. Para as composições alguns algoritmos são utilizados para padronização da imagem final.

3. Resultados

Para execução da tarefa, foi desenvolvido um *script* em Python que utiliza o TensorFlow para carregar o modelo treinado, executar a detecção e exportar as máscaras binárias dos objetos geradas pelo modelo. Um exemplo do resultado utilizando essa abordagem é apresentado na Figura 1.

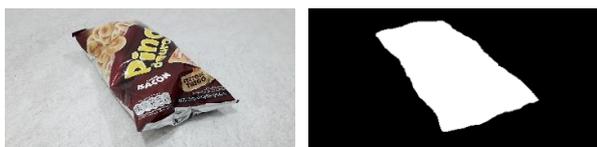


Figura 1 - Imagem do objeto (esquerda) e máscara gerada (direita).

Além disso, o programa utilizado pela equipe para fazer a composição das imagens finais, realiza as composições de acordo com o número de vezes que cada imagem de objeto deve aparecer no conjunto de dados e, ao final do processo, exporta as imagens finais junto com um arquivo contendo as anotações para treinamento do modelo, que são compostas pelo nome da imagem, classe do objeto e a posição da caixa delimitadora que marca o objeto na imagem, definida pelas coordenadas dos pixels X_{min} , X_{max} , Y_{min} e Y_{max} . Onde X representa o vértice superior esquerdo da caixa e Y o vértice inferior direito. Um exemplo de uma imagem criada pela ferramenta pode ser visto na Figura 2.



Figura 2 - Imagem final gerada pela ferramenta

A composição semântica da imagem não tem influência no desempenho do modelo. A Rede utiliza os pixels da região do objeto como exemplo positivo e os pixels das demais regiões como exemplo negativo. Então, mesmo que para um humano a imagem final não tenha sentido, ela é capaz de ensinar a Rede a detectar os objetos de interesse em aplicações reais.

O método apresentado tem se demonstrado, até o momento, eficaz para a geração de um bom *dataset* no cenário das competições robóticas. Para se obter uma detecção eficiente para realizar as tarefas do robô de serviço da equipe RoboFEI, necessitava-se de um *dataset* com cerca de 400 exemplos. Nesse caso, as imagens eram

coletadas e anotadas manualmente e esse processo levava cerca de 12h corridas. Utilizando a geração sintética, coletam-se apenas 10 imagens de cada objeto, em um fundo homogêneo, e gera-se um *dataset* sintético contendo aproximadamente 3000 exemplos em cerca de 4h. Além disso, embora as imagens reais sejam melhores para o desempenho da rede, com essa quantidade de imagens sintéticas, foi possível obter um desempenho equivalente nas detecções nos testes realizados.

4. Conclusões

A ferramenta desenvolvida se demonstrou eficiente, uma vez que tempo necessário para o processo de geração do conjunto de dados foi reduzido significativamente. O projeto encontra-se em fase de desenvolvimento e, para o relatório final, planeja-se a realização de um conjunto de experimentos, como testes de desempenho, testes de comparação com modelos treinados em conjuntos de dados criados a partir de imagens reais e, em todos os testes, fazer uma análise do tempo gasto desde a coleta de dados até a aplicação em uma tarefa real.

Este projeto é relevante para qualquer aplicação que envolva aprendizado profundo e que requer uma grande quantidade de dados. A detecção de objetos tem ampla aplicação em diversos cenários do nosso mundo e, com esta ferramenta, espera-se contribuir com o desenvolvimento das pesquisas, facilitando o trabalho de criação dos conjuntos de dados.

5. Referências

- [1] AQUINO JUNIOR, P. T. et al. HERA: Home Environment Robot Assistant. II BRAHUR and III BRASERO, 2019.
- [2] NOGUES, F. C. et al. Object Detection using Domain Randomization and Generative Adversarial Refinement of Synthetic Images, 2018. Ascent Robotics, Inc. Japan.
- [3] DWIBEDI, D. et al. Cut, Paste and Learn: Surprisingly Easy Synthesis for Instance Detection, 2017. The Robotics Institute, Carnegie Mellon University.
- [4] HOU, Qibin et al. Deeply Supervised Salient Object Detection with Short Connections. CoRR, abs/1611.04849, 2016. arXiv: 1611.04849.
- [5] XIE, Saining; TU, Zhuowen. Holistically-Nested Edge Detection. CoRR, abs/1504.06375, 2015. arXiv: 1504.06375.
- [6] MARTÍN ABADI et al. TensorFlow: Large-Scale Machine Learning on Heterogeneous Systems, 2015.
- [7] WANG, Jingdong et al. Salient Object Detection: A Discriminative Regional Feature Integration Approach. International Journal of Computer Vision, v. 123, n. 2, p. 251–268, 2017.

Agradecimentos

À instituição Centro Universitário FEI pela realização das medidas, empréstimo de equipamentos e apoio ao Projeto de Competições Robóticas RoboCup@Home.

¹ Aluno de IC do Centro Universitário FEI. Projeto (PBIC 107/19) com vigência de 12/19 a 11/2020.